IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Autonomous Clothes Manipulation using a Hierarchical Vision Architecture

LI SUN[1,2], (Member, IEEE), GERARDO ARAGON-CAMARASA[1], (Member, IEEE), SIMON ROGERS[1], and J. PAUL SIEBERT[1], (Member, IEEE).

[1]School of Computing Science, University of Glasgow, Glasgow, G12 8RZ, UK (e-mail: Gerardo.AragonCamarasa, Simon.Rogers, Paul.Siebert@gla.ac.uk)
[2]Lincoln Centre for Autonomous Systems, University of Lincoln, Lincoln, LN6 7TS, UK (e-mail: lsun@lincoln.ac.uk)

Corresponding author: Li Sun (e-mail: lisunsir@gmail.com).

*This work was supported in part by European FP7 Strategic Research Project, CloPeMa; www.clopema.eu

ABSTRACT This paper presents a novel robot vision architecture for perceiving generic 3D clothes configurations. Our architecture is hierarchically structured, starting from low-level curvature features, to mid-level geometric shapes and topology descriptions, and finally high-level semantic surface descriptions. We demonstrate our robot vision architecture in a customised dual-arm industrial robot with our in-house developed stereo vision system, carrying out autonomous grasping and dual-arm flattening. The experimental results show the advanced effectiveness of the proposed dual-arm flattening using the stereo vision system compared to single-arm flattening using the widely-cited Kinect-like sensor as the baseline. In addition, the proposed grasping approach achieves satisfactory performance on grasping various kind of garments, verifying the capability of the proposed visual perception architecture to be adapted to more than one clothing manipulation tasks.

INDEX TERMS Robot Clothes Manipulation, Visual Perception, Garment Flattening, Garment Grasping, Dual-arm Manipulation

I. INTRODUCTION

THE increasing need to deploy robots over a broader range of perception and manipulation tasks requires increased robot capabilities to compensate unpredictable settings and scenarios. Modern robots can perform a wide range of isolated tasks with high-precision, accuracy and reliability given that the environment and the task are rigid and static. However, robot perception and manipulation of deformable objects represents a difficult tasks for robots to perform consistently and accurately. This is because a robot needs to fully perceive and understand the state of a garment configuration at a given time during the manipulation task, and deformable objects can possibly take almost infinite configurations and shapes.

We hypothesised that a continuous sense-plan-act loop could indeed be exploited to overcome the limitations of the above non-trivial tasks [1]. In this paper, we, therefore, describe a novel robot vision architecture capable of perceiving and understanding deformable objects. Our architecture transforms low-level 3D visual features into rich semantic descriptions to underpin dexterous manipulation. We demonstrate that our robotic architecture can be employed to carry out the robotic tasks of autonomous dexterous grasping and dual-arm flattening.

Current research efforts have been advocated to solve subtasks within an autonomous laundering pipeline, these are: grasping clothes from a heap of garments [2], [3], recognising clothes categories [2], [4]–[7], unfolding [4], [8]–[12], garment pose estimation [13]–[16], garment flattening [8], [17], [18], ironing [19] and folding [20]–[24]. We found that existing approaches for garment perception have been devised as ad-hoc robot vision solutions rather than generalisable robot vision architectures that can be adapted into different robotic clothes manipulation tasks. Most simplify the perception task while mainly focusing on the manipulation aspects of the robotic task.

We aim to address the visual perception problem by significantly increasing the robot capabilities while handling deformable objects. To investigate this problem, we focused on developing robotic solutions for garment grasping and dual-arm flattening using our robot vision architecture. We choose these application areas because garment flattening has been under-developed in the literature thus far (with exception to our early work in [17] and [18]) and garment grasping provides a point of reference to benchmark our proposed vision architecture with respect to the state-of-art.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

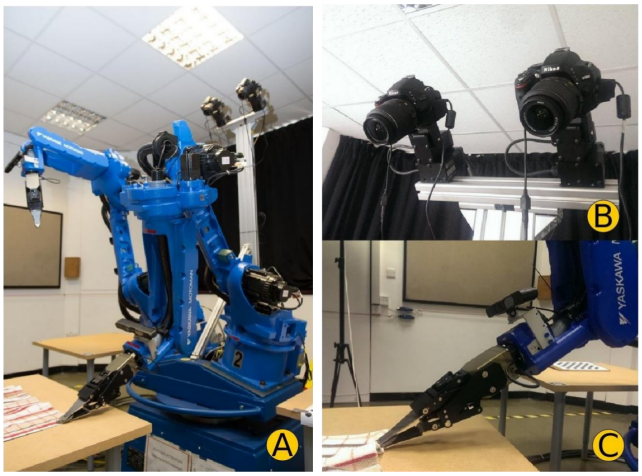


FIGURE 1. (A) The CloPeMa robot which consists of two six degrees of freedom YASKAWA arms and a custom made YASKAWA turn-table. Each arm features a specialised gripper for handling clothing and an ASUS Xtion Pro attached to the wrist of the arm. (B) Our stereo robot head integrated on our dual-arm robot. (C) A close up of the robot's gripper.

The main contributions of this paper are:

- Most of the existing approaches are devised for specific tasks, which lack a full-understanding of generic clothes configurations. In this paper, we propose a vision architecture with hierarchical 3D features for more than one manipulation tasks (i.e. clothes grasping and dual-arm flattening.
- The widely-used Kinect-like cameras (i.e. Kinect and Xtion pro) is not precise enough for dexterous clothes manipulation manipulation tasks e.g. flattening. Alternatively, we proposed an active binocular camera system for high-resolution depth sensing.
- To the best of our knowledge, our proposed dual-arm clothes flattening is the first and unique autonomous dual-arm flattening solution.

The structure of this paper is: In Section II, a comprehensive literature review presenting the state-of-the-art achievements of visually-guided garment manipulation. Section III provides an overall overview of our autonomous clothes manipulation system including our customised robot, stereo robot head, the hand-eye calibration and the proposed vision architecture. In Section IV, the visual architecture for generic garment surface analysis is detailed. Section V presents the proposed visually-guided grasping approach and dual-arm flattening approach. The experimental validations of the proposed autonomous grasping and flattening are detailed in Section VI. The conclusion of this work is given in Section VII.

II. LITERATURE REVIEW

Research on deformable objects in robotics is often framed to applications handling clothing items. The current generation robotic cloth perception and manipulation is introduced in five categories (according to the core tasks in autonomous laundering), namely, clothes grasping, unfolding, folding, flattening and the generic interactive perception. A summary of state-of-the-art approaches is thus given below, followed by a discussion and limitation on current state-art-of-the-art approaches.

a: Garment grasping

Ramisa et al. [2], [3] proposed a grasping detection approach using RGBD data. Their approach consisted of extracting SIFT and GDH (Geodesic Distance Histogram) local features in the RGB and depth domain, respectively, to detect wrinkled regions. After Bag-of-Features coding, two layers of SVM classifiers are trained with linear and χ^2 kernels. During the testing phase, a sliding window method is employed to detect graspable positions. After detection, 'wrinkledness' is calculated from the surface normals to select the best grasping location.

b: Garment Unfolding

The critical step for garment unfolding is to detect grasping locations that can potentially lead to an unfolding state (e.g. corners of a towel, shoulders of a shirt, waistline of a pant, and so forth). In this case, Cusumano et al. [8] have proposed a multi-view based detection approach for unfolding towels. Their technique is based on finding two corners that are along the same edge of a towel. Following on Cusamano's work, Willimon et al. [4] proposed an interactive perception-based strategy to unfold a towel on the table. Their approach relies on detecting depth discontinuities on corners of towels. For each iteration, the highest depth corner on the towel is grasped and pulled away from its centre of mass. This approach is constrained to a specific shape of cloth (square towel); hence it is unlikely to be extended to other clothing shapes.

Doumanoglou et al. [9], [10] have proposed a general unfolding approach for all categories of clothes. Their approach is based on active random forests and hough forests which are used to detect grasping positions on hanging garments. Unfolding is carried out by iteratively grasping the lowest point of the observed garment until an unfolding state is detected. In their later work [12], geometry-based visual clues, e.g. edges extracted from depth maps, are employed to detect the grasping positions for unfolding. Li et al. [11] have also devised an interactive unfolding strategy, in which the relevance of grasping positions for unfolding is modelled based on Gaussian density functions.

c: Garment Folding

Miller et al. [25] modelled each category of clothes with a parametric polygonal model. They proposed an optimisation approach to approximate polygonal models based on 2D contours on clothing obtained after segmenting the garment from the background. After that, the authors exploited their approach to fold garments based on a gravity-based [21] and geometry-based [22] folding strategies. In Stria et al.'s method [23], contour keypoints (i.e. collar points, sleeve

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

points, and so forth) are matched to a polygonal model stored in the database, thereby accelerating the matching procedure. In [26], the unfolding [9] and folding [23] are integrated.

d: Interactive Perception in Garment Manipulation

Interactive perception has been a critical role in dexterous clothing manipulation. That is, a robot iteratively changes the state of the garment from an unrecognisable or initial state towards a recognisable state. The working assumption of these approaches is that for each perception, there is a planned action; closely following a perception-manipulation loop. Specifically, Willimon et al. [27] first proposed to recognise the clothing's category from hanging configurations. In their approach, the hanging garment is interactively observed as it is rotated. In Cusumano et al. [8], the robot is driven to hang the garment and slid along the table edges (on both left and right side) iteratively until the robot can recognise its configuration and then to an unfolded configuration. In Doumanoglou et al.'s unfolding work [10], an active forest is employed to rotate the hanging garment to a perceptually-confident field of view. Li et al. [11] proposed a more straight-forward unfolding approach based on pose estimation [15], [16] by interactively moving the grasping point towards specific target positions (e.g. elbows). Moreover, interactive perception has been used in heuristic-based generic clothing manipulation.

In Willimon et al. [4] and our previous work [17], [18], perception-manipulation loops are carried out to track the flattening state of the garment and heuristic manipulation strategies are used to flatten the wrinkled garment on the working table.

e: Garment Flattening

For the specific garment flattening robotic task tackled in this paper, current research can be broadly classified into *gravity-based flattening* [9], [10] i.e. hanging the garment for reducing the wrinkledness and *by sliding the garment along a table* [8]. Since no visual-feedback is used in these methods, the garment cannot be guaranteed to be flattened.

Li, et. al [19] proposed a clothes ironing approach, in which multiple light sources are used to detect wrinkled regions for ironing. In our early research [17], [18], we first proposed to flatten wrinkled clothes on the table autonomously through detecting wrinkles and apply an heuristic to generate robot flattening actions on the detected wrinkles. Compared to the clustering-based methods (i.e. GMM used in [19] and K-means used in [17], a geometry-based method (i.e. used in [18]) results in a stable and precise visual framework for wrinkle detection and quantification.

A. DISCUSSION

The rigid objects (usually kitchen objects e.g. bottles, bowels, plates, cooking tools, etc.) are relatively smaller than garments, and therefore the grasping detection task can be addressed by exploring different grasping poses centred by the objects. However, the grasping detection or pose estimation-based methods [28]–[34] cannot be directly adapted to deformable garment grasping. The difficulty is that, more delicate configuration parsing and dexterous grasping are required in order to fetch small landmarks on garments (e.g. wrinkles, collars, cuffs, etc.). The main limitation of deep learning-based grasping detection [30], [32] and end-to-end motor-control methods [35], [36] is that large-scale training examples/trials are required. Although simulated objects can be utilised [32], [36], limited progress is achieved on deformable clothes due to the difficulty of simulating their physis.

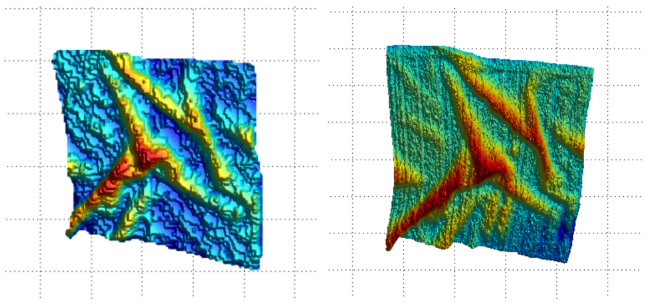


FIGURE 2. The comparison between depth data produced by Kinect-like camera and stereo head. (a) the depth map captured by ASUS Xtion Pro. (b) the depth map captured by stereo head.

The reported literature in deformable object perception and perceptions shares commonalities in the chosen RGBD sensor. Specifically, Kinect-like cameras are widely-used for perception tasks [2], [3], [9]–[11], [15], [16], [24]. Kinect-like cameras can provide real-time depth sensing with a precision of approximately 0.3-3cm[1] by trading off image resolution and depth sensing accuracy. Kinect-like cameras cannot capture small landmarks or estimate the magnitude of bending of clothes surfaces accurately, which is essential for successful dexterous manipulation as demonstrated in this paper. To this end, we integrated our custom-made active binocular robot head with our in-house 3D matcher (described in Sections III-B and III-D)

Reported methods are constrained to a specific garment or task at hand. In other words, current robot vision approaches for clothes manipulation are not generic enough for more than one task. Arguably, the latter can be attributed to the lack of sufficient understanding and perception of the clothes configuration, by which the generic landmarks can be localised and parametrised. Ramisa et al. [37] proposed a 3D descriptor that is employed for clothes grasping, wrinkle detection, and category recognition tasks. To the best of our knowledge, this descriptor is the only generic approach for multiple clothes perception tasks. However, Ramisa et al. evaluated their manipulation performance in annotated datasets as opposed to real-life experiments where camera

[1]This depth sensing precision depends on the range between camera and object. In our robotic scenario, the precision is about 0.5-1.0cm.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

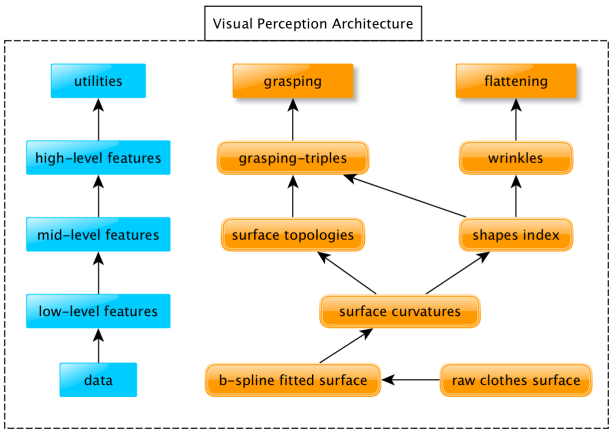


FIGURE 3. A hierarchical visual architecture for visually-guided clothes manipulation.

and robot calibration, mechanical and sensing errors and mistakes caused by labelling were not considered.

We can conclude that current approaches for dexterous clothes perception and manipulation have the following limitations. Firstly, Kinect-like and low-resolution depth cameras are not precise enough to sense garment details; hence, dexterous visually-guided manipulation become challenging for the task at hand, and complex heuristics need to be put in place. These types of cameras, therefore, constrain the application scope and capabilities of robots. Secondly, existing approaches usually focus on specific robot vision solutions rather than developing a general purpose robot vision architecture that provides a robot with the required abilities to understand surface shapes and topologies of deformable objects. As a consequence, most of the existing approaches are unlikely to be extended outside the application and task focus. We thus propose a robot vision architecture for dexterous clothing manipulation to advance the state-of-the-art in deformable objects perception and manipulation research.

III. AN OVERALL SCHEMA OF THE AUTONOMOUS SYSTEM

In this section, we introduce our proposed autonomous system for clothing perception and manipulation. This system consists of our customized dual-arm robot, our active binocular robot head and its calibration, stereo matcher, vision architecture, visually-guided manipulation skills and robot motion control.

A. CLOPEMA ROBOT

The main robot manipulators are based on the industrial robotic components for welding operation which are supplied by YASKAWA Motoman. As shown in Fig. 1-A, two MA1400 manipulators are used as two robot arms. Each manipulator is of 6 DOF, 4 kg maximal load weight, 1434 mm maximal reaching distance, ± 0.08 mm accuracy. These specifications satisfy the requirements for conducting accurate and dexterous clothing manipulation. They are mounted on rotatable turning tables. The robot arms and turning table are powered and controlled by a DX100 controller. The aim of CloPeMa project is to design a clothes folding prototype robot from (mainly) off-the-shelf components. We choose the YASKAWA arms because the size and the load of the MA1400 manipulators is capable of manipulating adult clothes. The approaches and methods developed in this paper are robot agnostic as all algorithms have been developed following ROS principles. Robotic manipulation and grasping are driven through the *MoveIt* library[2] such that it can be easily interfaced into another bi-manual robot configuration supported by ROS and *MoveIt*.

B. CLOPEMA ROBOT HEAD

Differing from most of the state-of-the-art visually-guided manipulation research, we aims to use relatively inexpensive, commercially available component elements to build an robot vision system (binocular head) for garment depth sensing.

In order to offset the limitation of widely-used depth sensor such as Kinect w.r.t. accuracy and resolution, a self-designed robot head is used in this paper for depth sensing. As shown in Fig. 1-B, the robot head comprises two Nikon DSLR cameras (D5100) that are able to capture images of 16 mega pixels through USB control. Gphoto library[3] is employed to drive the capturing under ubuntu. These are mounted on two pan and tilt units (PTU-D46) with their corresponding controllers. The cameras are separated by a pre-defined baseline for optimal stereo capturing. Its field of view covers the robot work-space. The robot head provides the robot system with high resolution 3D point cloud.

C. STEREO HEAD CALIBRATION

Our stereo head calibration has two steps: camera calibration and hand-eye calibration. The former is employed to estimate the intrinsic parameters of the stereo cameras. For the CloPeMa robot, the OpenCV calibration routines[4] are employed to estimate the intrinsic camera parameters of each camera. Furthermore, hand-eye calibration is employed to link the stereo head's reference frame into the robot kinematic chain. In other words, the unknown transformation from the camera frame to the calibration pattern coordinate system, as well as the transformation from the calibration pattern coordinate system to the hand coordinate system, need to be estimated simultaneously. For the CloPeMa stereo head, Tsai's hand-eye calibration [38], [39] routines are used to estimate rigid geometric transformations between camera to chess board and chess board to the gripper.

D. STEREO MATCHER

Having calibrated and integrated the stereo-head, the next stage is stereo-matching and 3D reconstruction. In this proce-

[2] Available in ROS: http://moveit.ros.org
[3] http://gphoto.sourceforge.net/
[4] http//opencv.org

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

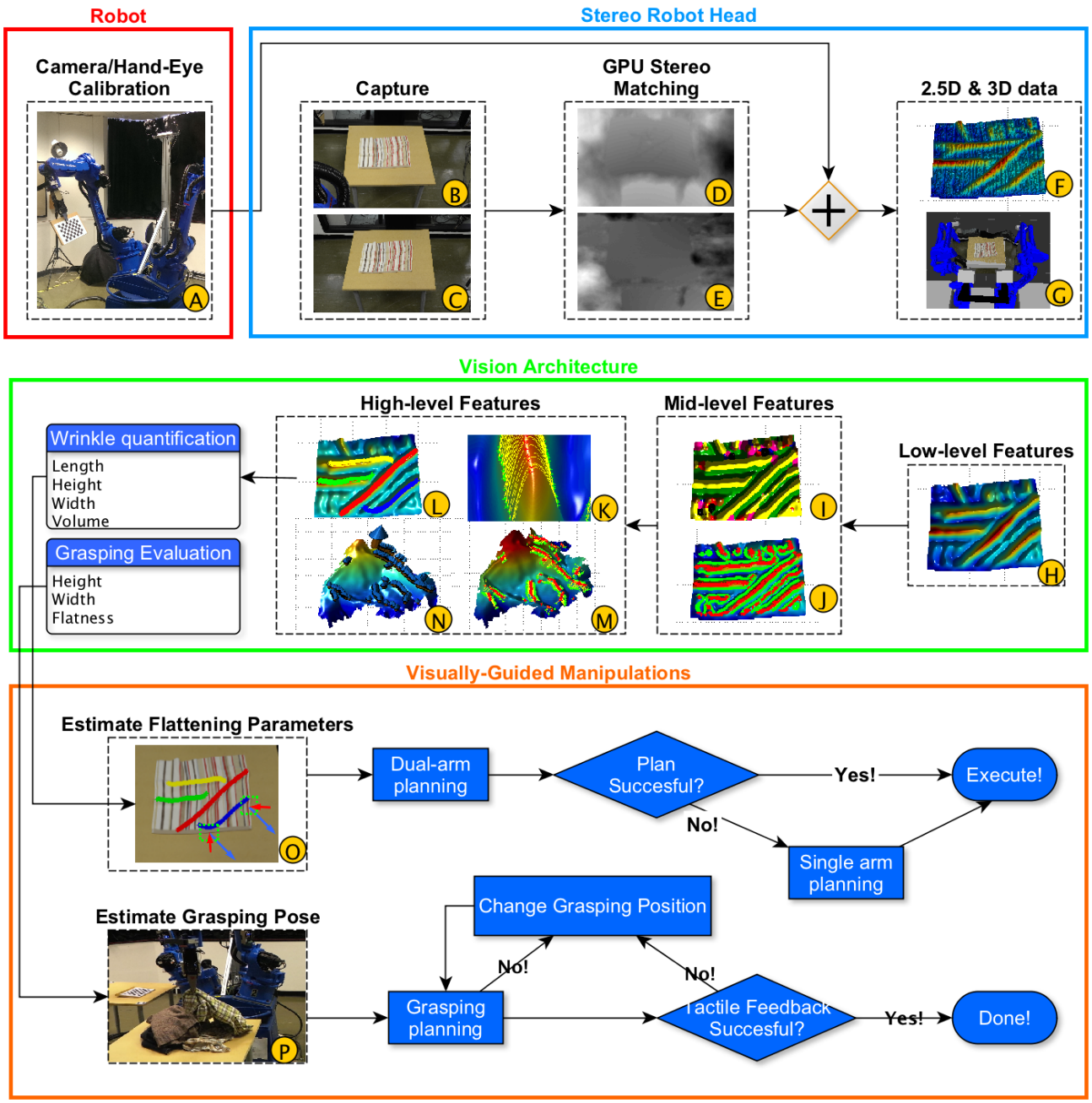


FIGURE 4. The whole pipeline for autonomous grasping and flattening.

dure, a pair of images are captured simultaneously by the left and right cameras. The C3D matcher [40], [41] is employed to find the horizontal and vertical disparities of the two images. In the implementation for the CloPeMa robot head, C3D matcher is accelerated by CUDA[5] GPU paralleling programming [42] to produce a 16 mega-pixel depth map in 0.2 fps. A GMM-based grab-cut [43] pre-trained by table color information is employed to detect and segment the garment.

E. ROBOT MOTION CONTROL

The ClopeMa robot is fully integrated with Robot Operating System through ROS industrial package[6]. More specifically, the *URDF* (uniform robot description form) is used to define the geometric structure of the robot. After the geometric

[5]https://developer.nvidia.com/cuda-zone

[6]http://wiki.ros.org/Industrial

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

structure is defined, collision can be detected by robot collision models, and the transforms between robot links can be achieved by TF[7]. *MoveIt* package is employed to achieve the communication between user interface and robot controllers.

F. THE CLOPEMA PROJECT AND ITS RELATED RESEARCH

During the project, tactile sensing, visual sensing and soft materials manipulation were jointly managed by a goal driven, high-level reasoning module. Inspired by the perception-manipulation cycle of the mammalian brain, the reasoning module also provided perception capabilities to fuse sensing and manipulation. The task calls for hierarchical representations and related perception-manipulation skills of different complexities. These theories addressed real-life autonomous laundering problems, e.g. dual-arm garment folding [23], unfolding [9], [10], dual-arm flattening [18], interactive sorting [6] and a novel gripper design [44].

It is worth noting that, in our previous research [6], [18], the proposed lower-level curvature feature and mid-level shape and topology feature are used for clothes category recognition. More specifically, in [18], the B-Spline patches are extracted to descrip the landmarks of clothes, and the histogram representations of shape index and topologies are extracted to represent the stiffness of the clothes' fabric attributes. Later, an interactive perception approach is proposed [6], where two manipulation skills, i.e. grasping-shake-drop, grasping-flip, are proposed to interact with the garment and maximize the visual perception confidence.

IV. HIERARCHICAL VISION ARCHITECTURE

This section presents the proposed vision architecture in details. Firstly, a piece-wise B-Spline surface fitting is adapted as pre-processing in Section IV-A, and the low-level feature extraction is presented in Section IV-B. In Section IV-C, surface shapes and topologies are introduced as the mid-level features. Afterwards, two high-level features i.e grasping triplets and wrinkle description, are reported in Section IV-D and Section IV-E.

A. PRE-PROCESSING: B-SPLINE SURFACE FITTING

As geometry-based 2.5D features such as curvatures and shape index are extremely sensitive to high frequency noise, a piece-wise B-Spline surface approximation is used to fit a continuous implicit surface onto the original depth map. More details are presented in our previous work [45].

B. LOW-LEVEL FEATURE: SURFACE CURVATURES ESTIMATION

To compute curvatures from depth, 2.5D points in the depth map (i.e. x, y and depth $-$ x and y are in pixels while depth is in metres) are examined pixel by pixel in order to find if they are the positive extrema along the maximal curvature direction. That is, given a depth map I, for each point p in I, the mean curvature C_m^p and Gaussian curvature C_g^p are firstly calculated by Eq. 1 and Eq. 2, where first derivatives f_x^p, f_y^p, and second derivatives f_{xx}^p, f_{yy}^p,f_{xy}^p are estimated by Gaussian convolution. Then, the maximal curvature k_{max} and minimal curvature k_{min} can be calculated by C_m^p and C_g^p (shown in Eq. 3).

$$C_m^p = \frac{(1+(f_y^p)^2)f_{xx}^p + (1+(f_x^p)^2)f_{yy}^p - 2f_x^p f_y^p f_{xy}^p}{2(\sqrt{1+(f_x^p)^2+(f_y^p)^2})^3} \quad (1)$$

$$C_g^p = \frac{f_{xx}^p f_{yy}^p - (f_{xy}^p)^2}{(1+(f_x^p)^2+(f_y^p)^2)^2} \quad (2)$$

$$k_{max}^p, k_{min}^p = C_m^p \pm \sqrt{(C_m^p)^2 - C_g^p} \quad (3)$$

C. MID-LEVEL FEATURES: SURFACE SHAPES AND TOPOLOGIES

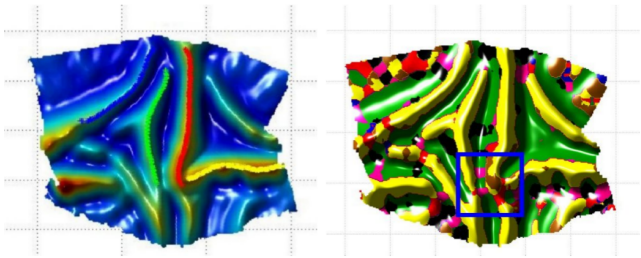
1) Surface Shape Analysis

Shape index [46], performs a continuous classification of the local shape within a surface regions into real-value index values, in the range [-1,1]. Given a shape index map S, the shape index value S^p of point p can be calculated as follows [46]:

$$S^p = \frac{2}{\pi}\tan^{-1}\left[\frac{k_{min}^p + k_{max}^p}{k_{min}^p - k_{max}^p}\right], \quad (4)$$

where k_{min}^p, k_{max}^p are the minimal and maximal curvatures at point p computed using Eq.3. The shape index value is quantised into nine uniform intervals corresponding to nine surface types – *cup*, *trough*, *rut*, *saddle rut*, *saddle*, *saddle ridge*, *ridge*, *dome* and *cap*.

In order to parse the shape information exhibited by the visible cloth surface, the shape index map is calculated from each pixel of the depth map and a majority rank filtering is applied. This non-linear filtering removes outlier surface classifications and can be tuned to produce a relatively clean classification of shape types over the cloth surface. An example can be seen in Fig. 4-I. It is worth noting that, the shape types 'rut' and 'dome' can be used to recognise the junction of multiple wrinkles thereby splitting wrinkles (as shown in Fig. 5).



(a) (b)

FIGURE 5. An example of splitting wrinkle using Shape Index. In highly wrinkled situations, the shape of wrinkles at junctions are classified as dome or rut (as shown in brown and red colours); this classification is used to separate jointed wrinkles in this work.

[7] http://wiki.ros.org/tf

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

2) Surface Topologies Analysis

Among all the shape types, *ridges* is of critical importance in the analysis and description of wrinkles. In this paper, the definition of 'ridges' shares similarities to that given by Ohtake et al. [47]. The main difference is that instead of estimating curvatures from a polygon mesh, surface curvatures are calculated using differential geometry, obtained directly from the depth map (as it is presented in Eq. 3). The *surface ridges* are therefore the positive extrema of maximal curvature while *the wrinkle's contour* is the boundary of the concave and convex surfaces of the garment.

From the nine shape types, four are convex (i.e. saddle ridge, ridge, dome, cap) and the rest are concave (i.e. cup, trough, rut, saddle rut, saddle). Thereby, the wrinkle's contour can be estimated. Alternatively, the boundary of the convex and concave surface can be more robustly estimated by computing the zero-crossing of the second derivatives of the garment's surface. In our implementation, a Laplace template window of size 16×16 is applied on the depth map in order to calculate the second order derivative. After the wrinkle's contour has been detected, the garment surface topologies are fully parsed. An example can be seen in Fig. 4-J.

D. HIGH-LEVEL FEATURES - GRASPING TRIPLETS

In this paper, a wrinkle comprises a continuous ridge line localised within in a region where the surface shape type is 'ridge'. The wrinkle is delimited (bounded) by two contour lines, each located on either side of the maximal curvature direction. A wrinkle can be quantified by means of a triplet comprising a ridge point and the two wrinkle contour points located on either side of the ridge, along the maximal curvature direction (as shown in Fig. 4-K).

The above definition is inspired by classical geometric approaches for parsing 2.5D surface shapes and topologies (i.e. shape index [46], surface ridges and wrinkle's contour lines). In this work, the height and width of a wrinkle are measured in terms of triplets. Accordingly, triplets can also be used as the atomic structures for finding and selecting grasping points (shown in Fig. 4-N).

1) Triplets Matching

From wrinkle's geometric definition, the maximal curvature direction θ can be calculated by Eq. 5.

$$\theta = \tan^{-1} \frac{\partial y}{\partial x}. \tag{5}$$

In this equation, ∂y and ∂x are the derivatives of k_{max}, computed by Gaussian convolution.

Given a ridge point p_r in a depth map I with scale φ_{L1}, this proposed method searches for the two corresponding contour points (p_c^l and p_c^r) over the two directions defined by θ and its symmetric direction using a depth based gravity-decent strategy. If the searched path is traversed in the same 'ridge' region as p_r (shown as yellow in Fig. 4-I), the process will continue. Otherwise, the searching will be terminated. Algorithmic details of triplet matching are described in our previous work [45].

Theoretically, every ridge point should be matched with its two corresponding wrinkle contour points. Whereas, due to occlusions and depth sensing errors, some wrinkle points fail to find their associated contour points and therefore do not generate a triplet. In order to eliminate the uncertainties caused by occlusions and errors, only triplets whose ridge points matched with both two wrinkle contour points are regarded as valid primitives for wrinkle quantification. An example of triplets matching is shown in Fig. 4-K and M. Given a triplet t_p containing one ridge point p_r and two wrinkle contour points p_c^1 and p_c^2, the height h_t and width w_t can be calculated from the embedded triangle (triplets) using Eq. 6. It is worth noting that, the triplet's points are transformed to the world coordinates, and as a consequence the unit of height h_t and width w_t is in meter.

$$\begin{aligned} h_t &= 2\,\frac{d\,(d-a)\,(d-b)\,(d-c)}{c} \\ w_t &= c, \end{aligned} \tag{6}$$

where $a = \parallel p_r, p_c^1 \parallel_2$, $b = \parallel p_r, p_c^2 \parallel_2$, $c = \parallel p_c^1, p_c^2 \parallel_2$, and $d = (a+b+c)/2$. The numerator of the right hand side of the equation is the area of a triangle embedded into the 3D space.

E. HIGH-LEVEL FEATURES: WRINKLE DESCRIPTION

1) Wrinkle Detection

The wrinkle detection consists of two steps: first, connecting ridge points into contiguous segments; second, grouping found segments into wrinkles (Fig. 4-L). More details are given in our previous work [45].

After wrinkles have been detected, for each wrinkle, a fifth order polynomial curve is fitted along its ridge points. A high order polynomial curve is adopted in order to ensure that it has sufficient flexibility to meet the configuration of the wrinkles (here fifth order works well in practice). The polynomial curve denotes the parametric description of a wrinkle, and the curve function is defined as:

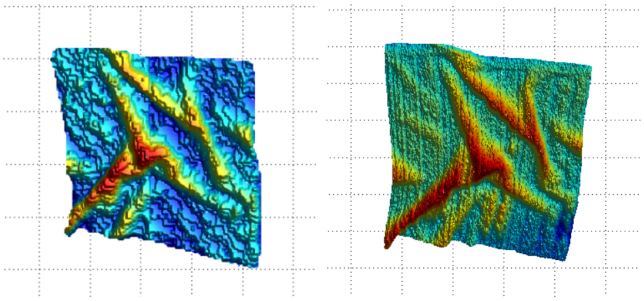
$$f(x) = (c_1, c_2, c_3, c_4, c_5, c_6) \cdot (x^5, x^4, x^3, x^2, x, 1)^T, \tag{7}$$

, here $c_1, c_2, c_3, c_4, c_5, c_6$ are the coefficients of fifth order polynomial curve.

2) Hough Transform-Based Wrinkle Splitting

As reported in Section IV-C1, *Shape Index* is used to find junctions of wrinkles, where the shape types 'rut' and 'dome' are used as the visual cues for splitting wrinkles (as shown in Fig. 5). Furthermore, an additional Hough transform-based wrinkle splitting approach is proposed. In our approach, the joined wrinkles are parametrised as straight lines in Hough space to find the primary directions of the joined wrinkles. Specifically, the Hough transform-based wrinkle splitting is used if the quality of wrinkle fitting (the *RMSE* of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access



(a) (b)

FIGURE 6. Splitting joined wrinkles through Hough-Transform based wrinkle direction analysis. (a) The Hough line parameter space. The x-axis refers to the angle in the Polar coordinate system α and the y-axis refers to the radius β. In this figure, the two peak points refer to the two main directions of the joined wrinkles. (b) The joined wrinkle points are split by two main directions. In this figure, the two main hough line directions are plotted as a blue line, and the points of joined wrinkles are split corresponding to their two directions, shown as red and green respectively.

polynomial curve fitting above) is not acceptable. That is, each 2D point on the fitted wrinkle is projected as a curve in the Hough parameter space. Afterwards, peaks in the Hough space are ranked and the two largest peaks indicate the two main directions of the joined wrinkles. In order to avoid choosing two peaks originating from the same wrinkle, the two largest peaks should satisfy a non-locality constraint. In the implementation of this work, the value of 20 degrees works well in practice. Then, wrinkle points can be split into two subsets depending on the two largest peaks. An example of the proposed splitting is shown in Fig. 6. Finally, new polynomial curves are approximated on the split points respectively. This splitting procedure will be performed recursively until all the wrinkles are below an optimal *RMSE* value (in practice, a value of 2 pixels works best for our implementation). Algorithm 1 details the proposed Hough Transform based wrinkle splitting approach.

3) Wrinkle Quantification

Shape Index classifies surface shapes without measuring surface magnitude. In our approach, the magnitude of a wrinkle's surface is measured by means of triplets (as described in Section IV-D1). Whereas, here the direction of triplet matching direction θ is estimated from the parametrised wrinkle description, which is more robust than that estimated from the maximal curvature direction. To be more specific, θ is computed from the perpendicular direction of the tangent line of the fitted curve on the observed wrinkle (i.e. fifth order polynomial curve in Eq. 7). The tangent direction δ can be calculated as:

$$\delta = \arctan((c_1, c_2, c_3, c_4, c_5) \cdot (x^4, x^3, x^2, x, 1)^T). \quad (8)$$

Having obtained the searching direction θ, the triplets on a detected wrinkle can be matched and thereafter their heights and width can be estimated by Eq. 6. That is, given a wrinkle, ω, containing a set of triplets $\{t_1, ..., t_{N_r}\}$, the width, w_w and

Algorithm 1 The Hough Transform based wrinkle splitting approach.

In: The detected wrinkles' points for splitting $\{P_x, P_y\}$, the threshold tolerance, $thres_{rmse}$, of the *RMSE* wrinkle fitting, and non-locality constraints threshold, $thres_\alpha$.
Out: The splited wrinkles' points $\{P_x^1, P_y^1\}$ and $\{P_x^2, P_y^2\}$.
Approximate polynomial curve to $\{P_x, P_y\}$, and calculate the fitting error $rmse$.
if $rmse$ is larger than $thres_{rmse}$ **then**
    Transform $\{P_x, P_y\}$ to Hough space, and get α and β in Polar coordinate system
    Find the peak points in hough space and rank them w.r.t the accumulator values $\{\hat{p}_1, ..., \hat{p}_{n_p}\}$.
    Find the two largest peak points $\hat{P}_1$ and $\hat{P}_2$ satisfying $\| \alpha_{\hat{P}_1}, \alpha_{\hat{P}_2} \| > thres_\alpha$.
    Restore two straight lines l_1 and l_2 in image space w.r.t two largest peaks in Hough space.
    Split the wrinkles' points $\{P_x, P_y\}$ into two subsets (P_x^1, P_y^1 and P_x^2, P_y^2) through calculating the minimal *Hausdorff* distances to l_1 and l_2.
else
    $\{P_x^1, P_y^1\}$ = $\{P_x, P_y\}$, and $\{P_x^2, P_y^2\}$ is empty.
end if
return $\{P_x^1, P_y^1\}$ and $\{P_x^2, P_y^2\}$.

height, h_w are calculated as the mean value of the width and height values of ω's triplets:

$$w_\omega = \frac{1}{N_t} \sum_{t_i \in \omega}^{N_r} w_{t_i}, \; h_\omega = \frac{1}{N_t} \sum_{t_i \in \omega}^{N_r} h_{t_i}, \quad (9)$$

where ω_{t_i} and h_{t_i} are the width and height of the ith triplet of the wrinkle ω.

For garment flattening, the physical volume of the wrinkle is adopted as the score for ranking detected wrinkles. PCA is applied on x-y plane of the largest wrinkle in order to infer the two grasping points and the flattening directions for each arm. More specifically, a 2 by 2 covariance matrix can be calculated from x and y coordinates, and then the principal direction of this wrinkle can be obtained by computing the eigenvector with respect to the largest eigenvalue. To obtain the magnitude that the dual-arm robot should pull in order to remove the selected wrinkle, the geodesic distance between the two contour points of each triplet are estimated. Section V-B details how these estimated parameters are used for flattening a garment.

V. GARMENT MANIPULATION USING THE PROPOSED VISUAL ARCHITECTURE

This section presents the autonomous robot clothes manipulation systems with integrated visual perception architecture.

The autonomous grasping approach is reported in Section V-A and dual-arm flattening is detailed in Section V-B.

A. HEURISTIC GARMENT GRASPING

In this research, two visually-guided heuristic grasping strategies are proposed, in which the high-level grasping triplet feature (Section IV-D) is adapted as the grasping location. Both strategies depend on an outlier removal strategy and grasping parametrisation for optimal garment manipulation as described in the following subsections.

1) Central Wrinkles Points Estimation

Due to stereo matching errors caused by occlusions, inaccurate and incorrect topological descriptions may be detected, thereby affecting the estimations of grasping candidates. A central point evaluation mechanism is therefore devised to deal with isolated and inaccurate detections. This mechanism consists of computing the *Mahalanobis* distance distribution of grasping triples. We adopt a *Mahalanobis* distancs based non-linear filtering. That is, given a grasping triplet t_i and the size of filter window[8], its *Mahalanobis* distance can be calculated as follows:

$$D_{Mahalanobis}(p_{t_i}, p_T) = \sqrt{(p_{t_i} - \mu_T)^T \Sigma^{-1} (p_{t_i} - \mu_T)} \tag{10}$$

, where p_{t_i} is the $x-y$ coordinate of t_i, T refers to all the triples within the filter window, μ_T is the mean of the $x-y$ coordinates of all triples, and Σ is the covariance matrix among all grasping triples within this region with respect to their spatial coordinates.

The probability of a grasping triplet being an outlier depends on the distance and direction within the spherical distribution. Hence, grasping triplets that are greater than a threshold[9] are treated as outliers and are removed from the list. This filtering is applied to every grasping triplet to probe whether it is an eligible grasping candidate.

2) Grasping Parameter Estimation

A *good grasping position* is considered as where the grasping region is most likely to fit the gripper's shape (constrained by the robot joints limitations) and at the same time is most unlikely to change the garment's configuration when grasped. That is, the gripper must get grip of a large region of the clothing surface in order to provide a firm grasp on the clothes. In this approach, two robotic poses are required for a successful grasping action. These are: *before-grasping* and *after-grasping* poses. The *before-grasping* pose is above the grasping point, while the *after-grasping* pose indicates the lowest position the gripper should reach without colliding with the surface of the garment. By interpolating these poses sequentially, the robot is therefore able to conduct a smooth grasping action.

The required parameters for completing the two grasping poses mentioned above comprise: the before-grasping pose of the gripper with respect to the robot's world reference frame, the normal vector of grasping triplet and the rotation angle of the gripper with respect to the normal vector. The 3D position of gripper can be indicated by the detected grasping candidate. The grasping orientation of the gripper is set as the surface normal direction of the local region to grasp. In this paper, the surface normals are robustly estimated from the third principal direction of PCA of local point cloud. In order to obtain a robust estimation of the gripper rotation, the principal direction of graspable candidates within a local region is estimated and its perpendicular direction is used as the gripper rotation.

3) Grasping Strategies

In this paper, two grasping strategies are proposed: a *height-priority* and a *flatness-priority* strategy. For the height-priority strategy, the grasping energy of the motion of the gripper is minimised by selecting the candidates from the highest graspable points with respect to the robot's world reference frame. While the flatness-priority strategy computes a flatness score for each grasping candidate, t, that encodes the height, h_t, and the width, w_t of the wrinkle's topology (Eq. 11):

$$flatness(t) = \frac{h_t}{w_t} \tag{11}$$

The height-priority strategy is able to grasp the clothing with the smallest cost of motion energy, and as a consequence the trajectory of planing is simpler, and is easier to solve the inverse kinematic problem and avoid collisions during motion planning. However, the drawback is also obvious, as the height-priority strategy cannot grantee that the mechanical shape of the gripper fits the region to grasp properly. In contrast, the flatness-priority strategy chooses the grasping candidate of the largest flatness rate, which is able to select the region most likely to fit the gripper but can bring difficulties to solving the inverse kinematic problem and avoiding collision. In our implementation, flatness-priority strategy and height-priority strategy are selected alternately until the grasping is completed.

B. DUAL-ARM GARMENT FLATTENING

1) Flattening Heuristic

In this research, the heuristic flattening strategy adopts a greedy search approach, in which the largest wrinkle detected is eliminated in each perception-manipulation iteration. Only largest detected wrinkle is considered to be modified per iteration because the manipulation errors accumulated into the system increases when considering a group of wrinkles with similar directions, and the likelihood of applying appropriate flattening is significantly reduced. Therefore, the largest wrinkle detection heuristic guarantees that a solution

[8]From practical experience, a filter window of 32×32 is used in our implementation.

[9]From practical experience, a threshold of 0.5 is chosen in our implementation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

is achieved regardless of highly wrinkled configurations. In our approach, wrinkles are quantified according to their physical volume in this chapter. The estimation of the volume of a wrinkle w is given by integrating the height of wrikle's surface points $h(u, v)$ on the two dimentional space u and v:

$$volume_w = \iint h(u,v)dudv \approx l_r * \frac{1}{N_t} \sum_{t_i \in w}^{N_t} (w_t \times h_t), \tag{12}$$

Practically, we approximate this integral by suming the uniform samples on the wrikle's surface. In this paper, we further simplify the approximation as the sum of uniform samples (i.e. 2D slice) on the ridge. Here N_t refers to the number of matched triplets, t_i is the ith triplet of w, w_t and h_t refers to the width and height of the triplet t_i, and l_r is the length of the wrinkle which is calculated by summing up the L^2 distances between every two nearest ridge points of the fitted wrikle.

2) Poses of a Primitive Flattening Action

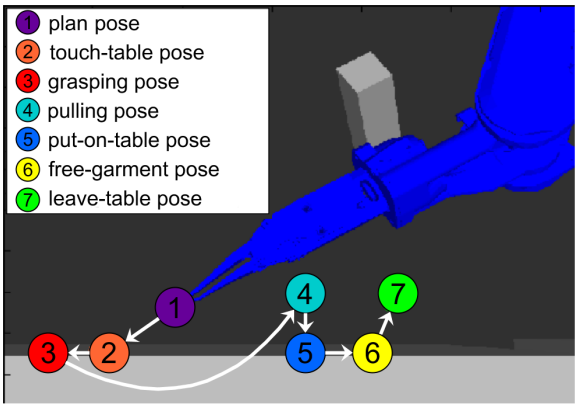


FIGURE 7. The seven poses for a robotic flattening motion. The gripper is moved to the 'plan pose', from where the trajectory of gripper is interpolated among poses sequentially in order to move the gripper. It is noticeable that the grasping direction and pulling direction are not aligned. The *plan pose*, *touch-table pose* and *grasping pose* are coplanar, while the *grasping pose*, *pulling pose*, *put-on-table pose*, *free-garment pose* and *leave-table pose* are coplanar. For the gripper state, it will be set to 'open' in *plan pose*, 'close' after *grasping pose* and 'open' again after *put-on-table pose*.

An entire flattening action consists of seven robotic poses: *plan*, *touch-table*, *grasping*, *pulling*, *put-on-table*, *free-garment* and *leave-table*. These poses are illustrated in Fig. 7. This figure also includes other pre-defined parameters used during the flattening task, e.g. orientation of the gripper w.r.t the table. The *plan pose* (Fig. 7, purple) refers to moving the robot's gripper close to the table by error-tolerance planing in preparation for flattening, then the gripper will approach the rest poses consequently by interpolating in *Cartesian* coordinates system. The *touch-table* and *grasping poses* (Fig. 7, orange and red, respectively) involve grasping the garment's boundary by interpolating the robot's motion between these two poses. The *pulling* and *put-on-table poses* pull the grasped garment according to the *Geodesic* distance

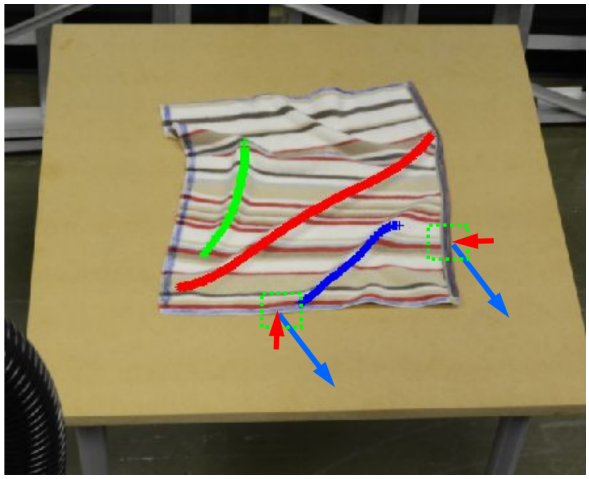


FIGURE 8. An example of detected wrinkles and the corresponding grasping poses and flattening directions of the dual-arms. The three largest wrinkles are shown, where the red one is the largest. The inferred grasping and flattening (pulling) directions are shown as red and blue arrows, respectively.

(Eq. 13) and smoothly return the garment to the table. Finally, the *free-garment* and *leave-table poses* are for freeing the garment and leaving the table.

In order to define these end-effector's poses and interpolate the trajectories, four parameters are required: *grasping position*, *grasping direction*, *flattening direction* and *flattening distance*. The *grasping* and *pulling poses* are estimated using these parameters. To be more specific, given the 3D grasping position and grasping direction (equivalent to the yaw rotation), the 6DOF pose of the gripper i.e. *grasping pose* can be obtained. The *pulling pose* can be then estimated via setting a proper flattening distance and direction. Afterwards, the other poses are inferred from the *grasping pose* and *pulling poses* by applying rigid transforms In this paper, these transforms are set according to the practical experience. For example, a translation of 4 cm along the grasping direction is set between *touch-table pose* and *grasping pose* to guarantee a firm but dexterous grasping of the clothes' edge. By interpolating these seven poses sequentially, the robot is therefore able to perform a smooth flattening action. It worth noting that for planning and interpolation, the *MoveIt* package is used. More details of the inference of four flattening parameters are given in the next section.

3) Flattening Parameters Estimation

Here the details about how to set the four flattening parameters (described in Section V-B2) are provided. As shown in Fig. 8, once the largest wrinkle is selected, *PCA* is employed to compute its primary direction, and the two *flattening directions* are perpendicular to the wrinkle′s primary direction. After the *flattening directions* are fixed, the two corresponding cross points on the garment contour are set as the position of the *grasping pose* (Fig. 7). In single arm flattening, the intersection point is defined between wrinkles bisector and garment's contour. Whereas, in dual-arm flattening, wrinkles

are divided into two equal segments and the two intersection points are calculated respectively. The *grasping direction* is estimated by the local contours of the *grasping positions* (as shown in Fig. 8). The pulling distance d_{w_i} of wrinkle w_i is estimated by:

$$d_{w_i} = \sum_{t_i \in w}^{N_r} (\mathrm{G}(c_l^{t_i}, c_r^{t_i}) - \mathrm{E}(c_l^{t_i}, c_r^{t_i}))/N_t) * Coeff_{spring}, \tag{13}$$

where t_i is the ith N_r triplet in w_i; $c_l^{t_i}$ and $c_r^{t_i}$ are its two wrinkle contour points; G refers to *Geodesic* distance [48][10], while E refers to *Euclidean* distance. $Coeff_{spring}$ is the maximal distance constraint between particles in a mass-spring cloth model[11].

4) The Dual-Arm Collaboration

Because of the limitation of the robot's joints and possible collisions between the two arms, not all of the *planing poses* (the first pose of a flattening action) can be planned successfully. Therefore a greedy *pose/motion exploration strategy* is proposed (The pseudo code of the proposed algorithm is provided in Algorithm 2). The goal of this algorithm is to explore the optimal pair poses of the two grippers and enhance the success rate of the motion planning of dual-arm flattening. More specifically, we define the error (offset in rotation) between the goal pose and the planning pose for both grippers i.e. e_l and e_r. Then we exhaustively explore all possible combinations of two grippersâĂŹ poses in a certain range with a proper interval and choose the pair poses with lowers joint error i.e. $e_l \times e_r$ to plan. This results in a significant improvement while flattening with both arms. However, if this algorithm fails, the robot only employs one arm; the arm used is selected according to the flattening direction[12].

VI. EXPERIMENTS

In this section, the proposed grasping and flattening approaches are evaluated in our dual-arm robot (Section VI-A and Section VI-B, respectively).

A. GARMENT GRASPING EXPERIMENTS

In this experiments, the grasping performance of the proposed approach is evaluated. The evaluation of robotic grasping has two parts: firstly, the success rate of single-shot grasping is investigated; secondly, the effectiveness of grasping is evaluated by counting the required number of shots for completing a successful grasping.

[10]Gabriel Peyre's toolbox is used in the implementation of this work for calculating *Geodesic* distance between two surface points: http//www.mathworks.co.uk/matlabcentral/fileexchange/6110-toolbox-fast-marching

[11]From practical experience, the $Coeff_{spring}$ is set as 1.10 in the experiments of this work.

[12]In order to enhance the success rate of motion planing, if the flattening action is towards left, then left arm is employed; otherwise, right arm is employed.

Algorithm 2 The Pose Exploration Algorithm for Planing Dual-Arms Grasping.

In: The direction interval is d_I. The maximum numbers of exploration in each side N_E.

Out: The final planable grasping directions of two arms d_L, d_R.

Compute the ideal grasping directions D_L, D_R.
if D_L, D_R is planable **then**
  $d_L = D_L$, $d_R = D_R$;
  return d_L, d_R
end if
Set the minimal whole error of two arms $e_{min} = \infty$
for $d_l = 0$; $d_l \leqslant d_I \times N_E$; $d_l = d_l + d_I$ **do**
  for $d_r = 0$; $d_r \leqslant d_I \times N_E$; $d_r = d_r + d_I$ **do**
    Compute the error of left arm and right arm, $e_l = d_l/d_I$, $e_r = d_r/d_I$;
    Compute the whole error $e_{lr} = e_l \times e_r$;
    if d_l, d_r is planable *and* $e_{lr} < e_{min}$ **then**
      $d_L = d_l$; $d_R = d_r$; $e_{min} = e_{lr}$;
    end if
  end for
end for

return d_L, d_R

1) Single-Shot Grasping Experiment

In the first grasping experiment, the grasping performance among five categories including t-shirts, shirts, sweaters, jeans and jackets are tested. Each category has three items of clothing, and 20 grasping experiments are tested on each item of clothing (in total 300 experiments). In each grasping experiment, the selected item of clothing is initialized to an arbitrary configuration by grasping and dropping it on the table. A successful grasping case means that: the gripper is moved to the position indicated by the visual feature; the grasping pose fits the shape of the region to grasp; and the clothing is fetched up. Since this work is focused on visual perception of grasping rather than kinematics, in these experiments, the flatness-priority grasping is carried out first. If the inverse-kinematics cannot be solved, the height-priority strategy is then applied (introduced in Section V-A3).

The experimental results of the first grasping experiment are shown in Table 1. Overall, the grasping success rate varies from 76.7% to 93.3% on different types of clothing. This difference can be attributed into the difference of clothes materials. In other words, the thickness and stiffness variation of clothes' materials brings different challenges to grasping. Specifically, the sweaters and t-shirts performed the best (93.3%,90%) while jeans and shirts obtained the lowest scores (78.3%,76.7%). The reason is two-fold: firstly, the more stiff the clothing material is, the more difficult the grasping is; and also, the more wrinkles the clothing configuration has, the easier the grasping is. On average, the proposed method is able to achieve 84.7% success rate

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

TABLE 1. The grasping success rate on different types of clothing.

Successful Rate	t-shirts	shirts	sweaters	jeans	jackets	average
categories	90.0%	78.3%	93.3%	76.7%	85.0%	84.7%
items	95%|85%|90%	70%|80%|85%	95%|95%|90%	70%|75%|85%	80%|95%|80%	–

TABLE 2. The required number of grasping shots for a successful grasping on different types of clothing.

Number of Shots	t-shirts	shirts	sweaters	jeans	jackets	average
categories	1.1	1.23	1.1	1.27	1.17	1.17
items	1.0|1.2|1.1	1.2|1.2|1.3	1.1|1.1|1.1	1.4|1.3|1.1	1.1|1.1| 1.3	–

among the five categories of clothing. In addition, the grasping performance on each item of clothing is also shown in the table. All 15 items of test clothing can achieve at least 70% successful grasping rate.

2) Multiple-Shot Grasping Experiment

Apart from the single-shot grasping success rate, the other criterion required to be evaluated is the number of trails for each completed grasping. The latter allowed us to demonstrate that our visual architecture and extracted features is able to handle difficult configurations[13]. In our implementation, the proposed grasping feature provides a ranked array of grasping candidates, then the robot attempts to grasp them sequentiality until the grasping is completed successfully. In order to acquire the grasping status, tactile sensors are used to detect whether the gripper is holding the garment.

Experimental results are shown in Table 2, in which 150 successful graspings are completed (10 experiments on each item of clothing) and 1.17 trails are required for each successful grasping on average. As shown in the table, similarly to the first grasping experiments, stiff clothes such as jeans and shirts require more grasping trails (1.27 and 1.23 times, respectively). The robot requires the least number of grasping trails on sweaters and t-shirts (1.1 and 1.1 times, respectively). The deviation between different items of clothing is small; the required number of trails ranges from 1.0 to 1.4 among all of the items of clothing. Among these 150 successful graspings, only 1 grasping is completed after 4 attempts, 3 graspings after 3 attempts, 17 graspings after 2 attempts, and the remaining 129 graspings are completed on the first attempt.

Overall, the experimental results of the proposed grasping method demonstrate a reliable grasping performance in terms of its grasping success rate (84.7%) and its effectiveness of grasping difficult configurations (1.17 trails on average).

B. GARMENT FLATTENING EXPERIMENTS

This section evaluates the performance of the proposed visual perception architecture on localising and quantifying wrinkles, and the integrated autonomous dual-arm flattening. This evaluation comprises three different experiments. Firstly, a benchmark flattening experiment comprising eight tasks is established to verify the performance and reliability for flattening a single wrinkle using dual-arm planning (Section VI-B1). While, in Section VI-B2, the second experiment demonstrates the performance of the proposed approach while flattening a highly wrinkled garment, comparing our robot stereo head system with standard Kinect-like cameras. Finally, Section VI-B3 investigates the adaptability of the proposed flattening approach on different types of clothing, in which the performance of flattening towels, t-shirts and shorts are evaluated and compared.

The proposed visual perception architecture is able to detect wrinkles that are barely discernible to human eyes unless close inspection on the garment is carried out. As it is not necessary to flatten these wrinkles, a halting criterion is therefore proposed, which scores the amount of 'flatness' based on the amount of the pulling distance computed in Eq. 13. In these experiments, if the flattening distances inferred by the detected wrinkles are less than 0.5 cm (barely perceptible), the garment is considered to be flattened[14].

1) Benchmark Flattening

The aim of the first experiment is to evaluate the performance of the proposed flattening method under pre-defined single wrinkle configurations as well as the dual-arm planning performance for flattening in different directions. For this purpose, eight benchmark flattening experiments are performed. As shown in Fig. 9, in each instance there is one salient wrinkle distributed in the range of 45 degrees to -45 degrees (from the robot's view). In order to obtain a stable evaluation, each experiment is repeated 5 times, and results are shown in Table 3.

It can be deduced from Table 3 that the proposed flattening approach is able to flatten these eight benchmark experiments with only one iteration. Moreover, the success rate for dual-arm planning is 85%, where the robot successfully grasps the edge(s) of the garment in all of these experiments. Experiment 5 shows a failed case while using both arms; this is because the robot reaches the limitation of its joints and the inverse kinematic planner adopted.

[13]Difficult configurations means those without graspable positions. They often appear in clothes made of stiff fabric; e.g. shirt and jeans.

[14]This value is obtained by averaging manually flattened garment examples performed by a human user.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

TABLE 3. The Required Number of Iterations (RNI) in the experiments.

Benchmark Experiments	exp1	exp2	exp3	exp4	exp5	exp6	exp7	exp8	average
RNI	1	1	1	1	1	1	1	1	1
Dual-Arm Success Rate	100%	100%	80%	100%	0%	100%	100%	100%	85%
Grasping Success Rate	100%	100%	100%	100%	100%	100%	100%	100%	100%

TABLE 4. The Required Number of Iterations (RNI) for flattening in highly wrinkled experiments. See text for a detailed description.

Flattening Experiments	exp1	exp2	exp3	exp4	exp5	exp6	exp7	exp8	exp9	exp10	AVE	STD	Dual-Arm Success
Dual-Arm (RH)	4(4)	5(4)	6(4)	5(4)	4(3)	5(3)	4(2)	5(2)	3(2)	6(3)	**4.7(3.1)**	**0.95**	**65.9%**
Dual-Arm (Xtion)	7(4)	8(4)	7(3)	12(4)	8(4)	13(7)	11(3)	10(5)	9(5)	10(5)	9.5(4.4)	2.07	46.3%
Single-Arm (RH)	7	12	5	8	7	7	12	14	8	6	8.6	2.99	-
Single-Arm (Xtion)	10	12	17	11	12	19	13	12	11	14	13.1	2.85	-

TABLE 5. The Required Numbers of Iterations (RNI) for flattening different types of garments.

RNI of tasks	exp1	exp2	exp3	exp4	exp5	exp6	exp7	exp8	exp9	exp10	AVE	STD
flattening towels	4	5	6	5	4	5	4	5	3	6	4.7	0.95
flattening t-shirt	5	7	12	11	7	8	12	9	12	6	8.9	2.68
flattening pants(shorts)	11	10	5	14	4	3	4	3	2	7	6.3	4.05

2) Highly-Wrinkled Towel Flattening

In order to investigate the contribution of the proposed dual-arm approach in terms of autonomous flattening of highly wrinkled garments, the flattening performance between single-arm and dual-arm strategies is compared. Similarly, in order to demonstrate the effectiveness of high-quality sensing capabilities for the dexterous clothes manipulation, a Kinect-like sensor is used as the baseline method (here the ASUS Xtion PRO[15] is used. We did not choose Kinect v1 or v2 due to the hardware configuration of our manipulator).

Therefore, for each experiment, a square towel is randomly wrinkled - wrinkles are distributed in different directions without following an order. Then different flattening strategies are applied (single-arm or dual-arm) with either the robot stereo head or Xtion. For comparison, 4 groups of experiments are carried out: (1) dual-arm using robot head, (2) single-arm using robot head, (3) dual-arm using Xtion and (4) single-arm using Xtion. To measure the overall performance and reliability, 10 experiments are conducted for each group of experiment and the required number of iterations (RNI) is counted as shown in Table 4. In Table 4, each column represents the experiment index for each of the groups proposed above. Values in parentheses show the RNI where dual-arm planning was successful while the rest of the values show the RNI for each experiment.

As shown in Table 4, the average RNI for dual-arm flattening using robot head is 4.7 (achieving 65.9% arm planning success rate) while single-arm is 8.6. The average RNI for dual-arm flattening using Xtion is 9.5 (achieving 46.3% dual-arm planning success rate), while single-arm is 13.1. This result shows that a dual-arm strategy achieves a much more efficient performance on flattening than a signal-arm strategy. The standard deviation (STD) of each group of experiments is also calculated, where the STD for dual-arm flattening is 0.95 (using robot head) and 2.07 (using Xtion) while for single-arm is 2.99 (using robot head) and 2.85 (using Xtion). As expected, a dual-arm strategy is not only more efficient but also more stable than a single-arm strategy. Likewise, from the sensors' perspective, the robot is able to complete a flattening task successfully within 4.7 iterations (dual-arm case) using the stereo robot head as opposed to 9.5 iterations while using Xtion. Overall, the our robot head clearly outperforms the Xtion in both dual-arm flattening and single-arm flattening experiments.

The results described above demonstrate that the dual-arm strategy is more efficient in flattening long wrinkles than the single-arm because the latter approach usually breaks long wrinkles into two short wrinkles. Likewise, comparing the our stereo head and Xtion, as observed during the experiments, it is difficult to quantify the wrinkles and also estimate the accurate flattening displacement (especially for small wrinkles) from the Xtion depth data because the depth map is noisier than the robot head (the high frequency noise is usually more than 0.5cm). Furthermore, long wrinkles captured by Xtion are often split into two small wrinkles due to the poor quality of the depth map, which in turn results in more flattening iterations (the short detected wrinkles are likely to have a lower dual-arm planing success rate).

3) Flattening Different Types of Garments

Since the proposed flattening approach has no constraints on the shape of the garment, this section evaluates the performance of this method on flattening other types of clothing, namely t-shirts and shorts. Ten flattening experiments are performed for each type of clothing. Examples are shown in Fig. 12 and Fig. 13, respectively.

The RNIs of different clothes categories are shown in

[15] $https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

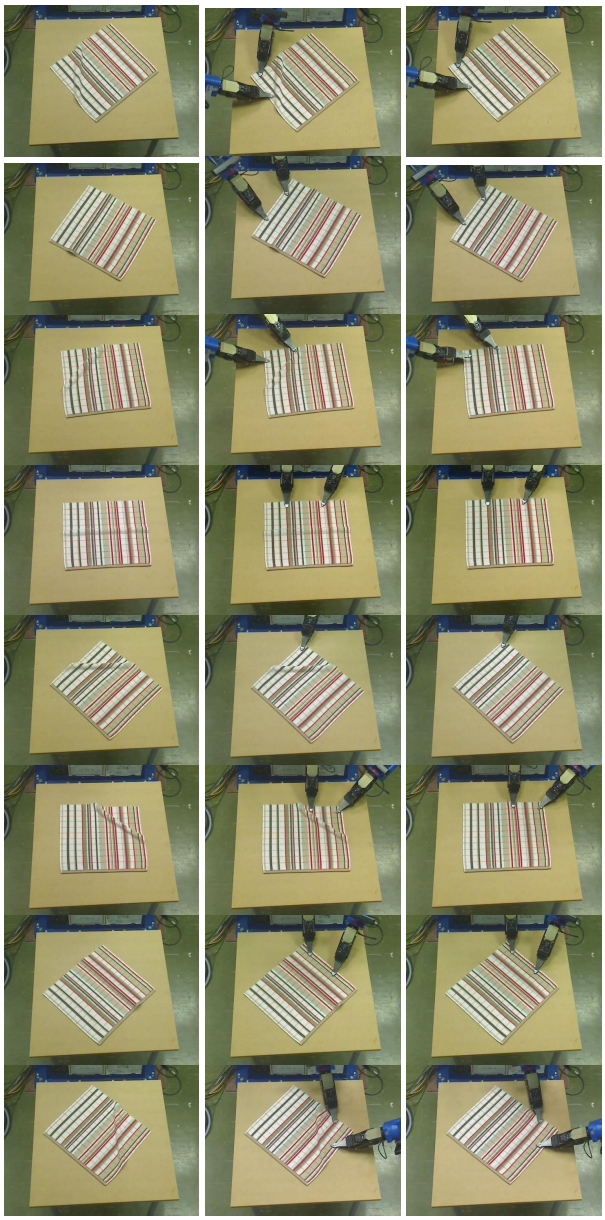


FIGURE 9. Eight benchmark experiments on a single wrinkle using dual-arm planning. Each row depicts an experiment, in which the left images show the stage before flattening; middle, during flattening; and right, after flattening.

Table 5, and here the towel flattening performance is presented as the baseline performance. As shown in the table, towels require an average of only 4.7 iterations to complete flattening. Shorts need more iterations on average(6.3) and t-shirts require still more (8.9). The reason is that towels are of the simplest shape among these three categories of clothing, while the shape of shorts is more complicated and that of t-shirts is the most complex. This experimental result demonstrates that the proposed approach is able to flatten different categories of clothing and that the RNI of flattening clothing is propagating to the complexities of the clothing's 2D topological shape.

C. SUMMARY

The proposed autonomous grasping is evaluated in both single-trial and interactive-trial experiments, showing robustness among the clothes types. And the validation of the reported autonomous flattening behaviours has been undertaken and has demonstrated that dual-arm flattening requires significantly fewer manipulation iterations than single-arm flattening. The experimental results also indicate that the dexterous clothes operation (such as flattening) is significantly influenced by the quality of the RGB-D sensor – using a customized off-the-shelf high-resolution stereo-head outperforms the commercial low-resolution Kinect-like cameras in terms of required number of flattening iterations (RNIs).

It takes approximately 50 seconds to grasp a clothing or apply a flattening iteration. This is mainly because we have to set the speed limit of the robot to 10% of the maximum speed for security reasons. Further more, considering the proposed vision architecture is not GPU parallelised, there exits a large room to improve the efficiency in the future.

VII. CONCLUSION

In this paper, a novel visual perception architecture is proposed for clothes configuration parsing, and this architecture is integrated with an active stereo vision system and dual-arm CloPeMa robot to demonstrate dexterous garment grasping and flattening. The proposed approach is based on generic 3D surface analysis, and tend to fully understand the landmark structures distributed on the clothing surface, thereby demonstrating the adaptability for multiple visually-guided clothes manipulation tasks. From the experimental validation, the conclusions are: firstly, the proposed visual perception architecture is able to parse the various garment configurations by detecting and quantifying structures i.e. grasping triplets and wrinkles; secondly, the stereo robot head used in this research outperforms Kinect-like depth sensors in terms of dexterous visually-guided garment manipulation; finally, the proposed dual-arm flattening strategy greatly improves garment manipulation efficiency as compared to the single-arm strategy. The integrated stereo head, visual perception architecture and visually-guided manipulation systems demonstrate the effectiveness of grasping and flattening different types of garments. On the other hand, the integrated autonomous flattening employs the perception-manipulation cycles, and consequently the clothing configuration is modified towards the flattening goal.

It is worth noting that our proposed vision architecture has the potential to be extended to more clothes perception and manipulation tasks. We has extend the proposed visual features to clothes recognition and sorting task [6], [7]. The future work will investigate the possibility of the proposed architecture for clothes on-table unfolding and folding. More invitigation will be done on depth sensing using stereo-based RGBD sensors e.g. Ensenso camera. Moreover, we use 16 megapixel image for stereo matching for a better accuracy and the vision architecture is implemented in CPU programming, thereby a close-loop manipulation with a real-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

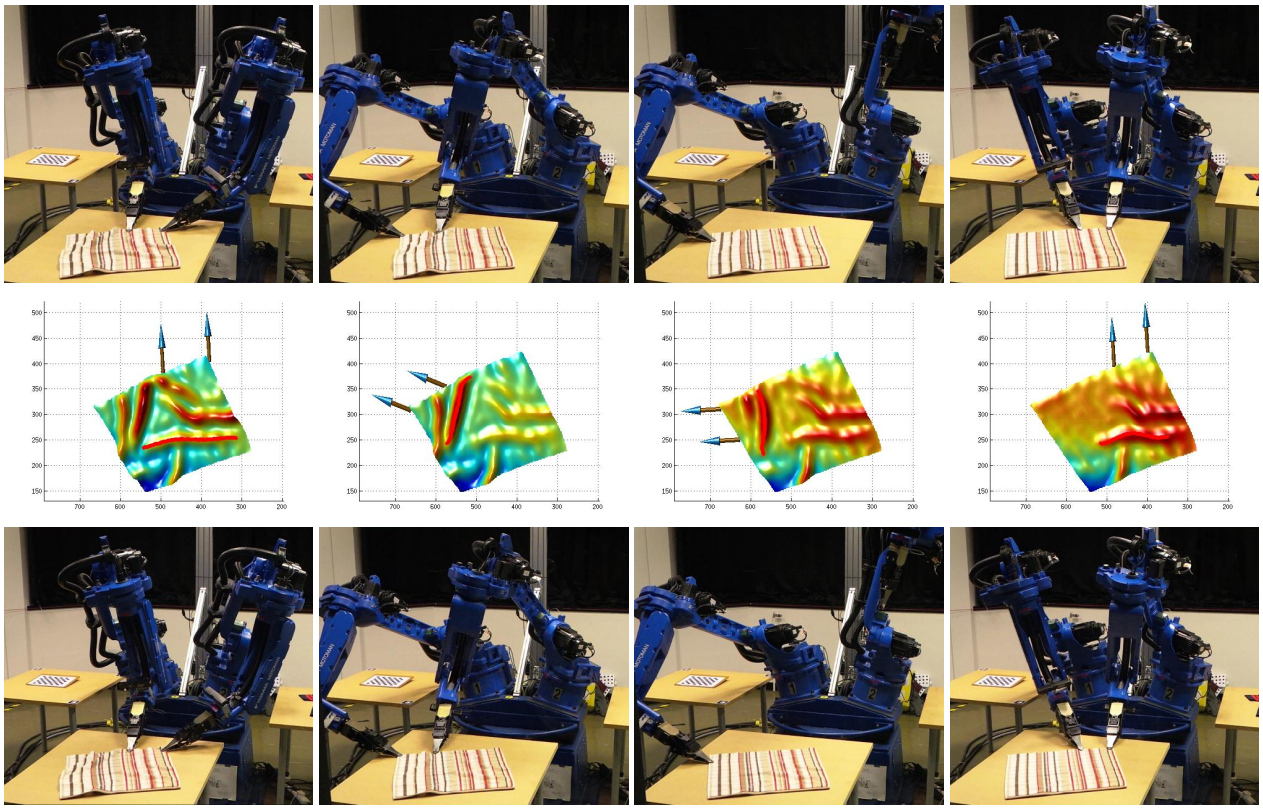


FIGURE 10. A demonstration of flattening an item of highly wrinkled towel. Each column depicts one iteration in the experiment. The top row depicts the towel state before the iteration; middle row, the detected largest wrinkles and the inferred forces; bottom row, the towel state after the iteration. On the third iteration, dual-arm planing demonstrated infeasible to execute, so a single-arm manoeuvre is formulated and applied.

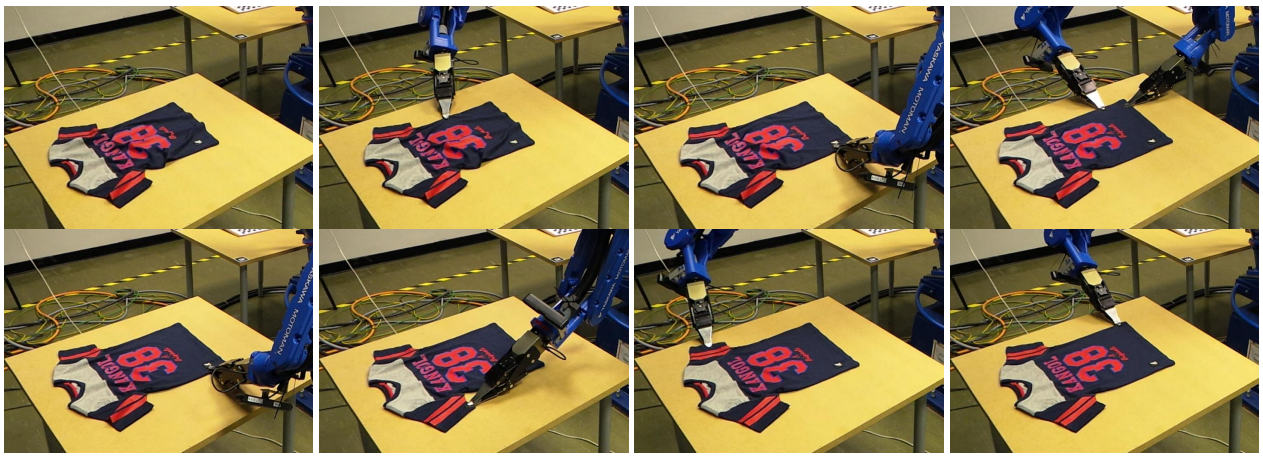


FIGURE 11. An example of flattening a T-shirt. As it is observed, the proposed flattening approach is able to adapt to any shape of garment, the robot can grasp the sleeves and stretch the wrinkles successfully.

FIGURE 12. Ten experiments of flattening t-shirts. Each column demonstrates a flattening experiment, in which the upper image refers to the initial configuration and the lower final configuration.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2883072, IEEE Access

FIGURE 13. Ten experiments of flattening shorts. Each column demonstrates a flattening experiment, in which the upper image refers to the initial configuration and the lower final configuration.

time perception is not achieved. In the future work, we will investigate the trade-off between performance and running time and reimplement the whole pipeline with GPU parallelisation.

ACKNOWLEDGMENT

We thank NVIDIA Corporation for donating a high-power GPU on which this work was performed. This project has received funding from the European Union's PF7 Specific targeted research projects (STREP) under grant agreement No 288553 (CloPeMa: Clothes Perception and Manipulation, http://www.clopema.eu/).

REFERENCES

[1] R. R. Murphy and A. Mali, "Lessons learned in integrating sensing into autonomous mobile robot architectures," Journal of Experimental & Theoretical Artificial Intelligence, vol. 9, no. 2-3, pp. 191–209, 1997.

[2] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras, "Using depth and appearance features for informed robot grasping of highly wrinkled clothes," in Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012, pp. 1703–1708.

[3] ——, "Finddd: A fast 3d descriptor to characterize textiles for robot manipulation," in Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, Nov 2013, pp. 824–830.

[4] B. Willimon, S. Birchfield, and I. D. Walker, "Model for unfolding laundry using interactive perception." in IROS, 2011, pp. 4871–4876.

[5] B. Willimon, I. Walker, and S. Birchfield, "A new approach to clothing classification using mid-level layers," in Robotics and Automation (ICRA), 2013 IEEE International Conference on, May 2013, pp. 4271–4278.

[6] S. Li, R. Simon, A.-C. Gerardo, and P. S. J., "Recognising the clothing categories from free-configuration using gaussian-process-based interactive perception," in 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016.

[7] L. Sun, G. Aragon-Camarasa, S. Rogers, R. Stolkin, and J. P. Siebert, "Single-shot clothing category recognition in free-configurations with application to autonomous clothes sorting," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sept 2017, pp. 6699–6706.

[8] M. Cusumano-Towner, A. Singh, S. Miller, J. F. O'Brien, and P. Abbeel, "Bringing clothing into desired configurations with limited perception," in Proceedings of IEEE International Conference on Robotics and Automation (ICRA) 2011, May 2011, pp. 1–8. [Online]. Available: http://graphics.berkeley.edu/papers/CusumanoTowner-BCD-2011-05/

[9] A. Doumanoglou, A. Kargakos, T.-K. Kim, and S. Malassiotis, "Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning," in Robotics and Automation (ICRA), 2014 IEEE International Conference on, May 2014, pp. 987–993.

[10] A. Doumanoglou, T.-K. Kim, X. Zhao, and S. Malassiotis, "Active random forests: An application to autonomous unfolding of clothes," in Computer Vision ECCV 2014, ser. Lecture Notes in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, vol. 8693, pp. 644–658.

[11] Y. Li, D. Xu, Y. Yue, Y. Wang, S.-F. Chang, E. Grinspun, and P. K. Allen, "Regrasping and unfolding of garments using predictive thin shell modeling," in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2015.

[12] D. Triantafyllou, I. Mariolis, A. Kargakos, S. Malassiotis, and N. Aspragathos, "A geometric approach to robotic unfolding of garments," Robotics and Autonomous Systems, vol. 75, pp. 233–243, 2016.

[13] Y. Kita, T. Ueshiba, E. S. Neo, and N. Kita, "Clothes state recognition using 3d observed data," in Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009, pp. 1220–1225.

[14] ——, "A method for handling a specific part of clothing by dual arms," in Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on. IEEE, 2009, pp. 4180–4185.

[15] Y. Li, C.-F. Chen, and P. K. Allen, "Recognition of deformable object category and pose," in Proceedings of the IEEE International Conference on Robotics and Automation, 2014.

[16] Y. Li, Y. Wang, M. Case, S.-F. Chang, and P. K. Allen, "Real-time pose estimation of deformable objects using a volumetric approach," in IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2014, pp. 1046–1052.

[17] L. Sun, G. Aragon-Camarasa, P. Cockshott, S. Rogers, and J. Paul, "A heuristic-based approach for flattening wrinkled clothes," in TAROS, 2013.

[18] S. Li, A.-C. Gerardo, R. Simon, and P. S. J., "Accurate garment surface analysis using an active stereo robot head with application to dual-arm flattening," in 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015.

[19] Y. Li, X. Hu, D. Xu, Y. Yue, E. Grinspun, and P. K. Allen, "Multi-sensor surface analysis for robotic ironing," in Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE, 2016, pp. 5670–5676.

[20] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding," in Robotics and Automation (ICRA), 2010 IEEE International Conference on. IEEE, 2010, pp. 2308–2315.

[21] J. Van Den Berg, S. Miller, K. Goldberg, and P. Abbeel, "Gravity-based robotic cloth folding," in Algorithmic Foundations of Robotics IX. Springer, 2011, pp. 409–424.

[22] S. Miller, J. Van Den Berg, M. Fritz, T. Darrell, K. Goldberg, and P. Abbeel, "A geometric approach to robotic laundry folding," The International Journal of Robotics Research, vol. 31, no. 2, pp. 249–267, 2012.

[23] J. Stria, D. Průša, V. Hlaváč, L. Wagner, V. Petrík, P. Krsek, and V. Smutný, "Garment perception and its folding using a dual-arm robot," in Proc. International Conference on Intelligent Robots and Systems (IROS). IEEE, 9 2014, pp. 61–67.

[24] Y. Li, Y. Yue, D. Xu, E. Grinspun, and P. K. Allen, "Folding deformable objects using predictive simulation and trajectory optimization," in Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on. IEEE, 2015, pp. 6000–6006.

[25] S. Miller, M. Fritz, T. Darrell, and P. Abbeel, "Parametrized shape models for clothing," in Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011, pp. 4861–4868.

[26] A. Doumanoglou, J. Stria, G. Peleka, I. Mariolis, V. Petrik, A. Kargakos, L. Wagner, V. Hlavac, T. Kim, and S. Malassiotis, "Folding clothes autonomously: A complete pipeline," IEEE Transactions on Robotics, vol. 32, no. 6, pp. 1461–1478, Dec 2016.

[27] B. Willimon, S. Birchfield, and I. Walker, "Classification of clothing using interactive perception," in Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011, pp. 1862–1868.

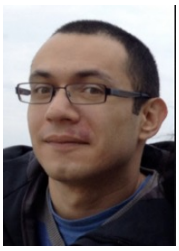
[28] A. Saxena, L. L. Wong, and A. Y. Ng, "Learning grasp strategies with partial shape information." in AAAI, vol. 3, no. 2, 2008, pp. 1491–1494.

[29] A. Saxena, J. Driemeyer, and A. Y. Ng, "Learning 3-d object orientation from images," in Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009, pp. 794–800.
[30] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," vol. 34, no. 4-5. SAGE Publications, 2015, pp. 705–724.
[31] J. Mahler, M. Matl, X. Liu, A. Li, D. Gealy, and K. Goldberg, "Dex-net 3.0: Computing robust robot suction grasp targets in point clouds using a new analytic model and deep learning," arXiv preprint arXiv:1709.06670, 2017.
[32] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," 2017.
[33] Y. Li, G. Wang, X. Ji, Y. Xiang, and D. Fox, "Deepim: Deep iterative matching for 6d pose estimation," in European Conference Computer Vision (ECCV), 2018.
[34] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," Robotics: Science and Systems (RSS), 2018.
[35] S. Levine, P. P. Sampedro, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," 2017. [Online]. Available: https://drive.google.com/open?id=0B0mFoBMu8f8BaHYzOXZMdzVOalU
[36] J. Mahler and K. Goldberg, "Learning deep policies for robot bin picking by simulating robust grasping sequences," in Proceedings of the 1st Annual Conference on Robot Learning, ser. Proceedings of Machine Learning Research, S. Levine, V. Vanhoucke, and K. Goldberg, Eds., vol. 78. PMLR, 13–15 Nov 2017, pp. 515–524.
[37] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras, "Finddd: A fast 3d descriptor to characterize textiles for robot manipulation," in IROS 2013. IEEE, 2013, pp. 824–830.
[38] R. Y. Tsai and R. K. Lenz, "Real time versatile robotics hand/eye calibration using 3d machine vision," in Robotics and Automation, 1988. Proceedings., 1988 IEEE International Conference on. IEEE, 1988, pp. 554–561.
[39] ——, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," Robotics and Automation, IEEE Transactions on, vol. 5, no. 3, pp. 345–358, 1989.
[40] J. Siebert and C. Urquhart, "C3d: a novel vision-based 3-d data acquisition system," in Image Processing for Broadcast and Video Production. Springer, 1995, pp. 170–180.
[41] J. Zhengping, "On the multi-scale iconic representation for low-level computer vision systems," Ph.D. dissertation, PhD thesis, the Turing Institute and the University of Strathclyde, 1988.
[42] P. Cockshott, S. Oehler, T. Xu, P. Siebert, and G. Aragon-Camarasa, "A parallel stereo vision algorithm," in Many-Core Applications Research Community Symposium 2012, 2012.
[43] J. Stria, D. Průša, and V. Hlaváč, "Polygonal models for clothing," in Proc. Towards Autonomous Robotic System (TAROS), ser. Lecture Notes in Artificial Intelligence, vol. 8717. Springer, 9 2014, pp. 173–184.
[44] M. J. Thuy-Hong-Loan Le, A. Landini, M. Zoppi, D. Zlatanov, and R. Molfino, "On the development of a specialized flexible gripper for garment handling," Journal of Automation and Control Engineering Vol, vol. 1, no. 3, 2013.
[45] L. Sun, S. Rogers, G. Aragon Camarasa, J. Siebert, and A. Khan, "A precise method for cloth configuration parsing applied to single-arm flattening," International Journal of Advanced Robotic Systems, 2016.
[46] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," Image and vision computing, vol. 10, no. 8, pp. 557–564, 1992.
[47] A. Belyaev and E. Anoshkina, "Detection of surface creases in range data," in Mathematics of Surfaces XI. Springer, 2005, pp. 50–61.
[48] J. A. Sethian, Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science. Cambridge university press, 1999, vol. 3.

LI SUN received the PhD from University of Glasgow in 2016. Now he is a post-doctoral research fellow with Oxford Robotics Institute, University of Oxford. He is IEEE, BMVA, EUcog, SICSA member.

From 2017 to 2018, he was a research associate with the Lincoln Centre for Autonomous Systems, University of Lincoln, UK. Before that, he was working as a research fellow at Extreme Robotics Lab, University of Birmingham, UK.

Dr. Li Sun's research focuses on the core challenges in the emerging robot vision to enable the robot to manipulate with complex industrial objects or drive in the dynamic, real-life environment e.g. warehouse, urban driving.



GERARDO ARAGON CAMARASA received the PhD from University of Glasgow in 2013. Now he is a Lecturer in Autonomous and Socially Intelligent Robotics in Computer Vision and Autonomous Systems group at the School of Computing Science, University of Glasgow.

From 2013-2016, he was working as a Research Fellow in Computer Vision and Graphics group, University of Glasgow and prior to that he did his PhD at University of Glasgow.

Dr. Gerardo Aragon Camarasa's research interests are in the multidisciplinary areas of robotics, chemical robotics, machine perception/vision and geometric algebras.

SIMON ROGERS received the PhD from University of Bristol in 2006. Now he is a senior lecturer in the School of Computing Science at the University of Glasgow.

Before starting a permanent post in Glasgow he did a couple of Post Doctoral positions (under Mark Girolami).

Dr. Simon Rogers's research involves the development of Machine Learning and Statistical techniques to help with the analysis of complex datasets, particularly within the field of Metabolomics but also other fields, Human-Computer Interaction and Information Retreival.

J .PAUL SIEBERT received his B.Sc. and Ph.D. degrees from the Department of Electronics and Electrical Engineering at the University of Glasgow, in 1979 and 1985, respectively. He is currently a Reader in the Department of Computing Science, University of Glasgow and the Computer Vision and Graphics group leader.

From 1991âĂŞ1997 he was with the Turing Institute, Glasgow, developing photogrammetry-based 3D imaging systems for clinical applications, and he served as Chief Executive from 1994. Prior to this he held the post of Scientist at BBN Laboratories, Edinburgh, from 1988 to 1991.

Dr. J. Paul Siebert's research interests include 3D imaging systems and tools for human and animal surface anatomy assessment, and also robot vision systems based on biologically motivated principles.